ZS Data Science Challenge – 2018

Singra Rupesh Duttatarya

Approach:

Before beginning of the actual coding a round of observation performed on the provide dataset. It was noticed that the training data had to merged into a singular file from three different files. When you go through the code you’ll notice first half of the coding is to modify dataset so that it output desired results. However the output file needed to presented in monthly detailed. Eventually the train data was also made into to same format.

Making of training data:

The train file which was provided was in week detail. It was later merged into monthly basis and expense details from expense file was added. Also holidays were counted on a monthly occurrence. It was also later added into the training file. All other columns like merchant details and S\_No were removed. All the above steps were repeated for test data as well. And later only necessary columns were kept.

Few details in the expense price file were missing. They were later made zero.